DOCUMENT RESUME

ED 069 092 EC 050 196

AUTHOR	Walker, Hill M.
TITLE	Early Identification and Assessment of Behaviorally Handicapped Children in the Primary Grades. Report No. 2.
INSTITUTION	Oregon Univ., Eugene. Dept. of Special Education.
SPONS AGENCY	Bureau of Education for the Handicapped (DHEW/OE), Washington, D.C.
BUREAU NO	BR-7-0706
PUB DATE	[71]
NOTE	67p.
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	*Behavior Problems; *Behavior Rating Scales; Classification; Emotionally Disturbed; *Exceptional Child Research; *Identification; Primary Grades; *Student Behavior

ABSTRACT

As part of a larger study investigating intervention procedures for children classified as homogeneous on factorially derived dimensions of classroom behavior, students in grades 1-3 (N=1,067) were screened using teacher ratings on the Walker Problem Behavior Identification Checklist (WPBIC) for the purpose of developing groupings of deviant classroom behavior using behavioral assessment procedures and factor analytic techniques. Each S's ratings on the WPBIC were scored on five factors and subjected to profile analysis. Homogeneous groupings were established on the five behavioral dimensions: acting-out, social withdrawal, distractability, disturbed peer relationships, and immaturity. Correlations indicated that, with the exception of acting-out and distractability, there was little overlap among item clusters comprising the five factors. Sex difference was significant within each of the three grade levels; neither grade level effect nor interaction between grade level and sex was significant. Results suggested that teacher checklist ratings of student behavior are a valuable and relatively inexpensive method of identifying homogeneous groupings of classroom behavior. (KW)

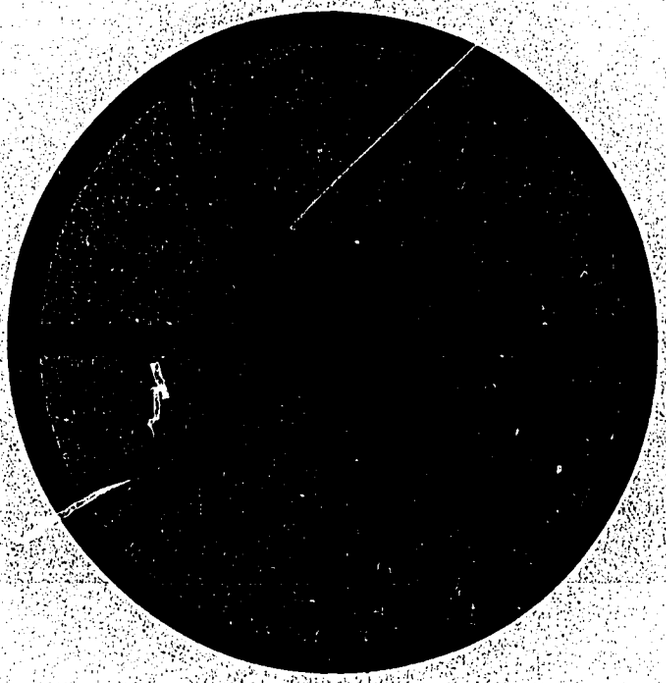
ED 069092

CORBEH

REPORT NO.2

EARLY IDENTIFICATION AND ASSESSMENT OF BEHAVIORALLY HANDICAPPED CHILDREN IN THE PRIMARY GRADES

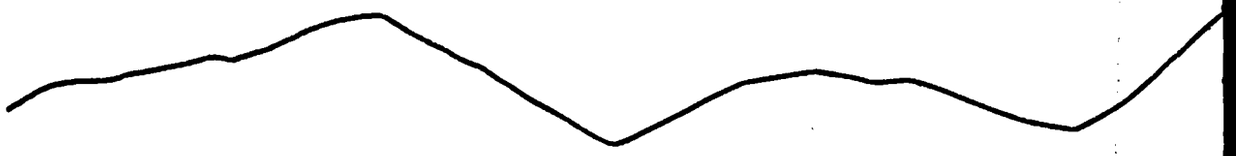
Hill M. Walker



Center at Oregon for Research in the Behavioral Education of the Handicapped, Department of Special Education, Clinical Services Building, University of Oregon, Eugene, Oregon 97403

EC 050 196 E

ED 069092



EARLY IDENTIFICATION AND ASSESSMENT OF BEHAVIORALLY HANDICAPPED CHILDREN IN THE PRIMARY GRADES

Report #2

Center at Oregon for Research in the
Behavioral Education of the Handicapped

The research or work report herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare through the Center at Oregon for Research in the Behavioral Education of the Handicapped, an R & D Center funded by the Division of Research, Bureau of the Handicapped.

Hill M. Walker
Department of Special Education
College of Education
University of Oregon
Eugene, Oregon 97403

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

Table of Contents

EARLY IDENTIFICATION AND ASSESSMENT OF BEHAVIORALLY HANDICAPPED CHILDREN IN THE PRIMARY GRADES

Hill M. Walker

University of Oregon

Department of Special Education

This report describes a normative identification study which had two primary objectives. These were: (1) to collect normative, behavioral assessment data on different subgroupings of deviant/deficient classroom behavior; and (2) to screen children in grades one, two, and three as possible candidates for a school-based intervention program designed to remediate classroom behavior problems.

Brief literature reviews are also provided in areas related to the identification of behavior problem children in the classroom setting. Areas reviewed include: (1) teacher ratings of classroom behavior, (2) early identification of behavior problem children, (3) factor analysis studies of classroom behavior dimensions, and (4) behavioral assessment and grouping for differential treatment.

The Teacher as an Observer of Classroom Behavior

The value and need for reliable identification of children with behavior problems seems to be generally accepted by educators and psychologists. However, since the publication of Wickman's 1928 monograph comparing the attitudes of teachers and clinicians toward behavior problems of children, the teacher's role in the identification process has been viewed with some equivocation. Wickman found discrepancy rank-order correlations of -.22 and -.11 respectively between the rankings of teachers and those of thirty mental health specialists on the relative seriousness of various problem behaviors of school children. Clinicians viewed social withdrawal and other anti-social forms of behavior as more serious, in terms of pathology, than did teachers. Teachers were more concerned with behaviors disruptive of classroom order, discipline, and achievement (Wickman, 1928). Since, in this study, the judgments of psychologists were accepted as a criterion against which teacher judgments were compared, the lack of agreement between these two groups raised serious questions about the competence of teachers in identifying disturbed children. On the other hand, Wickman's

research methodology has drawn considerable criticism that has cast some doubt upon the credibility of his findings. Watson (1933), for example, points out that teachers and clinicians were given different instructions for the rating/ranking task in the Wickman study. Teachers were instructed to rank behaviors for present seriousness, while clinicians were asked to rank them according to their importance or influence in handicapping a child's future adjustment.

Stouffer (1952) reported a study in which he used essentially the same research design as Wickman. His study demonstrated a much closer agreement, positive rho of .61, between teachers and mental hygienists in their ranking of the relative seriousness of children's behavior problems. In addition, Stouffer reported a rank order correlation of .87 between the ratings of his and Wickman's mental hygienists. Stouffer concluded that teachers' attitudes toward children's behavior problems had changed considerably since Wickman's study and had become more like those of psychologists.

Studies by Hunter (1957) and Ullman (1952) were also reported in the fifties, which showed greater congruence between teachers and mental health experts in their evaluations of childhood behavior problems than was true at the time of Wickman's study. Schrupp and Gjerde (1953) in a replication of Wickman's research design, found much more agreement between teachers and clinicians than was indicated in studies reported during the late 1920's. The authors qualified their findings by pointing out that disagreements were still evident, and that the direction of the disagreements was similar to that found by Wickman.

Different results were reported in studies by Clark (1951) and Peck (1955). Peck's study revealed that teachers viewed undesirable

personality traits as the most seriously handicapping of behaviors; regressive traits were slightly less serious; and aggressive behavior was rated as least serious. Clark concluded from the results of his study that teachers are more disturbed by children's behaviors which annoy other children than by behavior that affect teachers directly.

In the early sixties, Sarason (1960) and his associates maintained that developing personality measures to identify children whose anxiety levels are interfering with a productive use of their potential is important because teachers do not perform this function to a satisfactory degree. Sarason suggests teachers do not have either the time or the training to act as psychological diagnosticians.

Bower (1960) also used clinicians' judgments of emotionally disturbed children as a criterion variable to evaluate teachers' judgments of the same sample on dimensions of emotional disturbance. Bower found a very close relationship between teachers' and clinicians' judgments of emotional disturbance. In this study, teachers identified 87% of clinically identified children and identified a greater number of children as overly withdrawn or timid than as overly aggressive or defiant. Evidence from this study appears to refute the oft-cited criticism that teachers tend to ignore withdrawn children whose behavior may not be as disruptive or disturbing as that of an acting-out, aggressive child.

Beilin (1959) has reviewed research from 1927 that relates to the validity of teachers' identification of children with behavior problems. His interpretation of research findings suggests that teachers have become more like clinicians in making judgments about children. Beilin feels teachers and clinicians will likely always

differ in basic attitude. Teachers, because they are task-oriented, will probably focus more on problems disruptive of achievement than will clinicians. Clinicians, on the other hand, are traditionally more concerned with the child's adjustment. Thus, according to Beilin, they are more likely to identify withdrawn children even though they may be achieving satisfactorily (Kennedy, 1965).

Maes (1966) reported a study which demonstrates that emotionally disturbed children in grades four, five, and six can be identified as effectively through the use of a teacher rating scale and a group intelligence test as through the use of these two sources of information, in addition to arithmetic achievement, reading achievement, a modified sociometric technique (a class play), and a self-concept inventory. The predictive efficiency that Maes achieved with two variables (teacher ratings and intelligence estimates) equalled that which Bower demonstrated with the use of six variables. This procedure makes the identification process considerably more efficient and lends further support to Bower's finding that teacher judgment is an important variable in the identification of children with behavior problems.

Mathew Trippe (1961), in a discussion of the teacher's role in the identification of children with behavior problems, argues that competent teachers are the most qualified judges of disturbed behavior in the school setting. He notes that requiring the judgments of teachers to be validated against the judgments of clinicians fails to recognize the role of teaching as different from the clinician's role of diagnosing and administering treatments. Failure to distinguish between these roles has resulted in some concern that teachers might indiscriminately label children as deviant or disturbed. He suggests there is

no evidence to support this and that if a variety of school options were available, teachers' attention to children with behavior problems would result, not in the treatment of an illness, but in a better placement for the child.

Thus, evidence exists that teachers are in much closer agreement with mental health specialists in their judgments of classroom behavior problems than was true thirty years ago. Although some questions are still raised about the validity of teacher judgments of childhood adjustment problems, as well as the wisdom of using clinicians' and mental hygienists' judgments as a validation criterion, there appears to be a general recognition that the classroom teacher's vantage point is an especially good one for the initial stages of identifying such children. In fact, the classroom teacher is in a unique position to identify children with behavior problems, since she spends more time in actual observation of the child than any other school personnel (Kennedy, 1965).

A number of researchers have designed identification systems and procedures that rely heavily upon the teacher's judgment of classroom behavior problems. (Becker, 1960; Cromwell, 1965; Dreger, 1964; Ross, Lacey, and Parton, 1965; Bower, 1960; Quay and Quay, 1965; Zax, Cowen, Izzo, and Trust, 1964; Spivack and Swift, 1966; Swift and Spivack, 1968; Phillips, 1968; and Walker, 1969). Many of the identification instruments used in these studies consist of stimulus items that describe behaviors which interfere or actively compete with successful academic performance. According to Beilin (1959), teachers are most concerned with classroom behavior which is disruptive of achievement. Since the teacher is held responsible for the child's achievement through the teaching-learning process, she should be an excellent

judge of classroom behavior that is incompatible with academic performance.

The Case For Early Identification

The need for early identification of children with learning (Haring and Ridgway, 1967; and Fitzsimmons, Cheever, Leonard, and Macunovich, 1969) and behavioral (Robins, 1966; Bower, 1960; Cowen, Zax, Izzo, and Trost, 1966; O'Neal and Robbins, 1958; Zax and Cowen, 1967; Zax, Cowen, Izzo, Madonia, Merenda, and Trost, 1966; and Cobb, 1970) problems has received increasing attention in the last few years. Evidence from the above studies suggests that children with academic and behavioral difficulties can be identified early in their school careers. Fitzsimmons, Cheever, Leonard, and Macunovich (1969) analyzed the academic histories of 270 students from elementary through secondary school using pattern analysis and nonparametric techniques. Their analyses revealed that a majority of academically unsuccessful students (high school drop-outs--poorly performing graduates) could be identified as early as the third grade. The authors indicate that by the second grade, 50% of the 270 students in the sample had experienced their first academic failure. By the fourth grade, 75% had experience their first failure, and 90% by the seventh grade. The most critical areas of initial difficulty were in the basic skills subjects of language and mathematics.

Of greater significance was the finding that over 40% of the student records demonstrated a spread pattern (initially failing in only one or two academic areas). Distribution of the spread patterns through the school years showed a fairly consistent pattern beginning

with an initial increase in the first three years (usually in the basic skills area); then showing a more gradual rise over the years (spreading to science and social studies areas); and finally reaching a high point in the ninth and tenth grades. The spread effect indicates that many children give early warning signs of serious academic difficulties, and that they are likely to fall further and further behind in their academic skills the longer they remain in school. Although they present no data to support the hypothesis, the authors contend that the findings of their study suggest that intervention early in the child's academic career has more impact than later intervention. The authors reinforce their point by referring to the longitudinal research (not quoted or documented) which suggests that the further along in a student's career, the greater is the amount of "end-career" variance already accounted for.

They recommend special remedial programs for the first four grades in which children experiencing academic difficulties would be assisted in improving the quality of their achievement. They argue further that particular attention should be paid to the basic skills area to prevent a spread of performance difficulty to other areas.

In the area of behavioral disturbance, several studies have demonstrated correlational relationships between behavioral problems evidenced early in the child's school career and later maladjustment. Stennett (1965), for example, found that with the passage of time, school children identified as emotionally handicapped performed significantly less well than their peers. Westman, Rice, and Bermann (1968) reported a correlation of .88 between maladjustment ratings of

children early in their school careers and their subsequent use of mental health services (Zax, Cowen, Rappaport, Beach, and Laird, 1968).

Zax, Cowen, Rappaport, Beach, and Laird (1968) reported a study in which they identified two consecutive groups of first grade children having a high potential for being emotionally disturbed. Children manifesting high potential for being disturbed (labeled Reg Tag) were compared on school record and special test measures, with peers evidencing low potential for disturbance (labeled Non-Red Tag). The measures reflected achievement, classroom behavior, peer perceptions, attendance, and school nurse referrals. Forty-seven comparisons were made between the Red Tag and Non-Red Tag groups in the seventh grade. Ten of the 47 differences were statistically significant beyond the .05 level, which is a greater number than would be expected on a chance basis. All Red Tag children scored more negatively than the Non-Red Tag children on all significant differences. In addition, of the 37 non-significant differences, 30 found the Red Tag children scoring more negatively.

Forty-two comparisons were made between the second group of Red Tag and Non-Red Tag children identified in the first grade. Thirteen of the 42 differences were statistically significant, and the Red Tag group scored more negatively than the Non-Red Tag group on all significant comparisons. As with the first group, a majority of the non-significant comparisons (25 of 29) favored the Non-Red Tag group. The findings of this study have important implications for current special educational practices. The results suggest that children experiencing behavioral difficulties can be identified in the first grade. Further, the data indicate that behavior problems identified in the first year

remain stable over time. Problems identified in grade one have a high probability of being identified seven years later. The specific problem behaviors may change over time. However, evidence for the stability of behavior disturbance appears to be quite strong. The authors argue that, "... if, as seems valid on the face of it, children who manifest signs of poor adjustment are more likely than others to grow up to be seriously disturbed, then considerable effort at early identification of potential for maladjustment and the development of programs to prevent this are justified."

The work of Cobb (1970), in progress, on the identification and measurement of observable, achievement related behaviors in the first grade, is quite timely. His research design is sequential in that correlational relationships are established between predictor variables (observable behaviors) and a criterion of measured academic achievement. The identified, achievement related behaviors will then be modified across children to determine if functional relationships exist between them and academic achievement.

Thus, it appears a technology is developing that will allow the identification, prediction, and possible prevention of behavioral and academic difficulties in young children. Zax, Cowen, Rappaport, Beach, and Laird (1968) used an elaborate clinical procedure similar to Bower (1960), in identifying his groups of Red Tag children. They suggest the identification process in general needs further study as a source of information for the development of optimal prevention procedures. Additionally they argue that the identification process must be made more efficient and streamlined. Their procedure, as well as the observation system developed by Cobb (1970), is quite expensive in terms

of observer and teacher time. Both systems require training before they can be used effectively. It would appear that several levels of early screening in the school setting may prove functional as well as economical. Walker (1969) has described such a model in an earlier paper. It uses a 50 item behavior checklist as an initial screening device (requiring approximately five minutes to complete per child). High scoring children are then selected for more intensive screening and evaluation using direct observation and recording procedures.

Studies Using Factor Analysis and Clustering Techniques

A number of recent studies have factor analyzed ratings of child behavior by teachers, parents, and clinicians in an attempt to isolate homogenous behavioral groupings. The number of factors obtained in these studies has varied from two (Peterson, 1965) to as many as thirteen (Spivack and Swift, 1966). Peterson (1965), after reviewing a number of studies using child behavior scales, argues that two major factors account for the important variance in ratings of child behavior. The content of the first factor described by Peterson, relates to the behavioral dimensions underlying the child's social adjustment. The second factor describes behavioral dimensions associated with extroversion-introversion. Becker and Krug (1964) suggest that the type of factor analysis procedure used may determine the number of factors actually obtained. In ratings of child behavior, one typically finds two major centroid factors accounting for as much as half the variance, accompanied by a series of smaller factors. If, however, analytic rotational procedures are used (oblimax or varimax), Becker and Krug argue

that five to eight factors with reliable variance contribution are likely to be obtained. Thus, there appears to be a general lack of agreement among investigators about the number of dimensions that are necessary and sufficient to account for behavioral differences among children (Sines, Pauker, Sines, and Owen In Press). There is little doubt, however, that homogeneous groupings of ratings of child behavior can be identified and isolated (Becker and Krug, 1964; Patterson, 1964; Kulik, Stein and Sarbin, 1968; Ross, Lacey, and Parton, 1965; Sines, Pauker, Sines and Owen, In Press; Phillips, 1968; Quay, 1964; and Walker, 1970).

Patterson (1964), for example, factor analyzed clinic ratings of a sample of 100 boys between the ages of seven and twelve years. The analysis procedure yielded five factors which the author labeled as hyperactive, withdrawn, immature, aggressive, and anxious. Patterson set up a profile analysis procedure based upon the factor structure. The homogeneity of the obtained factor profiles were then analyzed. The hyperactive, withdrawn, and aggressive profile groups were the most homogeneous with intra class correlations respectively of .55, .63, .52. The immature and anxious groups were less homogeneous with coefficients of .42 and .39. All five factor profile groups were more homogeneous than a sixth group of subjects, labeled random, with an intra class R of .11.

Ross, Lacey, and Parton (1965) developed the Pittsburg Adjustment Survey Scales to provide for an objective evaluation of the social behavior of elementary school age boys, using the observations of classroom teachers. An initial item pool of 94 items was obtained through use of an extreme group procedure. Behavior ratings were

obtained on 209 boys in grades one through six. Each teacher in the sample rated one randomly selected boy in her class. A principal-components factor analysis of the data yielded five factors which accounted for 40% of the total variance or 71% of the estimated non-error variance. Factor V, which contained only one item with a loading in excess of .50, was dropped from the analysis. The remaining four factors were labeled aggressive behavior, withdrawn behavior, pro-social behavior, and passive-aggressive behavior. Additional analyses indicated that the factor scales discriminated among independently selected groups of aggressive, withdrawn, and well-adjusted school children. For example, a group of 18 aggressive boys received mean scores of 94.4 on aggressive behavior; 11.1 on withdrawn behavior; 0.0 on pro-social behavior; and 77.8 on passive-aggressive behavior. A group of 18 well-adjusted boys received mean scores of 5.6, 5.6, 33.3, and 11.1 respectively on the same factor scales.

Kulik, Stein, and Sarbin (1968) constructed a self-report checklist of antisocial activities for analyzing patterns of delinquent behavior. The study had three objectives: (1) to establish the dimensionality of adolescent antisocial behavior, (2) to identify salient patterns of antisocial behavior among consistently delinquent boys, and (3) to demonstrate validity of dimensional and pattern analyses by relating dimensions and patterns to other variables.

The 52 items of the checklist asked the subjects about a broad range of misbehaviors. Cluster analysis of the items on three different samples yielded four dimensions of antisocial behavior: delinquent role, drug usage, parental defiance, and assaultiveness. The checklist was

filled out by 505 high school boys and 391 boys at institutions for delinquents. The scores of delinquents and non-delinquents differed significantly on each of the four dimensions of antisocial behavior. Delinquent boys in the study were classified into seven empirical types based upon their score patterns on the four dimensions. The empirical types differed in racial composition and on other social and personal variables.

Quay has conducted a number of factor analytic studies of ratings and case histories of adolescents, children in special classes, and delinquent boys (Quay, 1964; Quay, Morse, and Cutler, 1966). Quay has identified four homogeneous factors or dimensions in these studies. They are inadequate-immature, neurotic-conflicted, unsocialized aggressive or psychopathic and socialized or sub-cultural delinquency. Quay points out that these behavior dimensions occur in delinquent, emotionally disturbed, and "normal" populations. Differences among these three groups on the four dimensions are quantitative rather than qualitative. The magnitude of the scores varies from sample to sample, but the dimensions remain the same (Quay, 1970).

Sines, Pauker, Sines and Owen (In Press) developed the Missouri Children's Behavior Checklist which provides a set of descriptions of children's behavior that may be rated by a child's parent. The purpose of the study was to develop a method ... "for identifying groups of children, each of which would be at the extreme of one of several clinically or theoretically significant dimensions of children's behavior." The final form of the checklist consisted of 70 statements that were reduced from 95 descriptive behavioral statements. The

original behavior statements were selected from the existing literature to sample six dimensions of behavior: aggression, inhibition, hyperactivity, sleep disturbance, somatization, and sociability. Items were assigned to behavior dimensions if a point beserial correlation between the item and the total dimension score was .30 or greater, and if the square of the point biserial r was at least twice as large as the square of the r between that item and the total score or any of the remaining five factors. This analysis was completed on parental ratings of 404 boys between the ages of five and sixteen years. The means and standard deviations, on each of the six behavior dimensions, were compared for 24 boys seen in a university child psychiatry clinic with a group of 24 non-referred boys who were evaluated and classified as "normal" children. There were statistically significant differences between the two groups of boys on the checklist scales of aggression, inhibition, hyperactivity, and sociability.

Walker (1970) factor analyzed behavior checklist ratings (by teachers) of 534 children in grades four, five, and six. Boys and girls were included in the sample. The procedure yielded five factors that were subjected to a varimax orthogonal rotation to obtain a simple structure. The five factors were: acting-out, withdrawal, distractability, disturbed peer relationships, and immaturity. Analyses revealed statistically significant differences in total checklist score between males and females across all three grade levels. Statistically significant differences in checklist score were found between a group of emotionally disturbed children and a matched group of non-disturbed children. A profile analysis procedure, based on the factor

structure, was established to record and analyze scores on each of the factor scales.

The factor analysis techniques employed in the above studies are useful in establishing the validity of an instrument, since they provide information about the content of a scale (what it measures). These procedures also provide for a more detailed description of behavior through factorial, profile analysis techniques. The factorial dimensions identified in the above studies share a high degree of similarity in number as well as content. The strongest and most homogeneous factors in these studies appear to be aggression, withdrawal, and hyperactivity (Patterson, 1964; Sines, Pauker, Sines, and Owen, In Press). Behavior dimensions associated with anxiety, immaturity, and disturbed peer relationships appear to be less well defined and less homogeneous, but still clearly identifiable. Several of these studies demonstrate that different clinically identified or independently selected groups of children received differential ratings on the factorial dimensions. Thus, powerful evidence exists in the literature for the identification of homogeneous groupings of deviant behavior, as well as for the external validity of such groupings.

Consequently, it would appear that children receiving high scores on different behavior dimensions can be grouped for the purpose of providing differential treatments. However, the basis for grouping and for assignment to a treatment rests upon a rating by a teacher, a parent, or a clinician as to whether or not a given behavior is present in a child's repertoire. If the child receives a large number of deviant behaviors checked or rated on a factor scale, he is said to

score high on, and be representative of, the behavior dimensions measured by that factor. Although such ratings are quite useful for identifying and locating specific populations of deviant children, they do not predict the actual rates with which these behaviors occur and therefore provide little information for the development of treatments and remediation procedures. There have been no studies reported in the literature which demonstrate that actual rates of individual behaviors can be predicted or inferred from checklist ratings of whether the behavior is present or absent. In addition, no study has demonstrated a relationship between the number of behaviors indicated as present on a checklist and the rate of occurrence of such behaviors as measured by direct observation and recording procedures.

It would appear that a homogeneous pool of subjects with respect to a given behavior dimension, such as hyperactivity or social withdrawal, would be highly variable in terms of the rate with which they produce the behaviors making up the behavior dimension. Some of the subjects would no doubt have very high rates; others moderate rates; and some low rates. Thus, in developing treatments for differential groupings of deviant behavior, it would seem necessary to also develop homogeneous groupings with respect to the rate with which individual behaviors comprising the behavior dimension occur. For example, in developing a treatment for social withdrawal, an initial group of subjects could be identified on the basis of high scores received on a factor scale within a checklist which measures social withdrawal. The next level of screening, prior to assignment to treatment, would require the identification of a pool of subjects, from the initial group, who

have low rates of social interaction. The second level of screening provides direct information for the development of intervention procedures. Similar screening procedures could be established for such factors as aggression, hyperactivity, deviant peer relationships, distractability, etc. using observation schedules. In summary, this model simply requires a more empirical definition of factorial homogeneity, and uses rate as a basis for assignment to treatment as well as for evaluating the effectiveness of intervention.

Assessment and Grouping for Differential Treatments

In the field of behavior modification, intervention procedures have traditionally been designed to shape or modify the behavior(s) of a single child. These single subject designs have focused upon precise analyses of the parameters of the target behavior(s) selected for modification. Intervention procedures have been adapted to the specific remediation requirements of the target behavior(s) as well as the reinforcement preferences of the child. Dunn (1968), for example, has pointed out that the intervention program itself often becomes the diagnostic device. The success of this individualized approach to assessment and remediation has been impressive. However, specific intervention programs across children have thus varied as a result of: the target behaviors selected for remediation, situational variables associated with different treatment settings, and specific remediation requirements and reinforcement preferences of different target children. As a result, the large number of individual case studies and single subject designs reported in the literature have not resulted in

clearly validated techniques or procedures that have a predictable effect across children or across behaviors. There is evidence in the literature that such techniques as social reinforcement, token reinforcement, and time-out procedures are effective in remediating behavioral deficits in specific instances. Nevertheless, there is no data to indicate under what treatment conditions, with what types of children, and across which behaviors are these techniques consistently effective in remediating behavior.

Increasing attention is being given to the development of "group" intervention techniques that can be used simultaneously with a large number of children and that will have some generality of effect both across children and across behaviors (Packard, 1970; Walker and Buckley, 1970; Walker, Mattson, and Buckley, In Press). It would appear that the effective education of behaviorally handicapped (as well as other types of handicapped children) requires the development and validation of intervention procedures that are effective; that have some generality of effect--both across children and across behaviors; that have some generality of effect over time; and that are reasonably economical in terms of per child cost.

Quay (1968) has provided a framework for delivery of remediation services to handicapped children that focuses upon assessment, grouping, and remediation. Quay's model is somewhat unique in that children with learning or behavioral handicaps are assessed on a variety of educationally relevant measures and then grouped for remediation and instruction according to their performance on these measures. Homogeneous groupings are established on dimensions of educationally relevant performance

instead of upon hypothetical medical or psycho-social correlates of handicapping conditions. Thus, homogeneous groupings are established for instructional-remedial purposes, across children and across handicapping conditions.

Purpose of the Study

The present study reports the behavioral assessment procedures and results for a larger study, the purpose of which was the development and evaluation of intervention procedures for children classified as homogeneous on factorially derived dimensions of classroom behavior. Specific objectives of the study are: (1) to develop homogeneous groupings of maladaptive or deviant classroom behavior using behavioral assessment procedures and factor analytic techniques; (2) to experiment with intervention strategies based upon the assessment data, that are specifically designed for remediation of behavioral deficits isolated by the grouping procedure; (3) to measure the efficiency and effectiveness of the intervention strategies in remediating behavioral deficits and producing behavior change.

Method

Assessment and Sample Selection Procedures

The population of children in grades one, two, and three in the Eugene school system was screened using teacher ratings on the Walker Problem Behavior Identification Checklist (WPBIC) (Western Psychological Services, 1970). The school district required parental permission for completion of the ratings. A checklist was completed on each child for whom a signed permission slip was received. Of 5,500 children in grades one, two, and three, parental permission slips were received and teacher ratings were completed for 1,067 children.

Children who received a checklist score of 21 or above--one standard deviation above the mean of the normative sample--were assigned to a pool for possible selection as experimental subjects. Each subject's ratings (by his teacher) on the WPBIC were scored on five factors within the checklist and subjected to a profile analysis procedure. Through this procedure, five pools of behaviorally homogeneous subjects were selected for further observation, screening and assessment.

Observation schedules will be developed to provide more precise and more reliable measurement of the behavioral content of each factor. The observation schedules will be based upon the behavioral content of each factor in the checklist. These schedules will provide observation and recording of discrete units of behavior within the classroom setting. (An observation schedule for factor one, acting-out behavior, has been developed and is included as appendix one.)

Each pool of behaviorally homogeneous subjects will be screened on the observation schedule developed for that factor. Subjects will then be drawn from this pool and assigned to an intervention procedure designed to remediate behaviors measured by that particular factor.*

The observation schedules will serve three functions in this study: (1) checking and corroboration of the teacher's ratings of classroom behavior on the WPBIC; (2) providing additional measures of factorially homogeneous behaviors through observation and recording of discrete behavioral units; (3) providing a basis for evaluating the efficiency and effectiveness of experimental intervention procedures. Five pools of homogeneous subjects were established on the following behavioral dimensions: (1) Acting-out (disruptive, aggressive, defiant); (2) Social Withdrawal (restricted

*The larger research study will last five years. One year will be devoted to developing intervention procedures for each of the five groups.

functioning, avoidance behavior, low rates of peer interaction); (3) Distractability (short attention span, inadequate study skills, high rates of non-attending); (4) Disturbed Peer Relationships (inadequate social skills, high rates of coercive manding, high rates of dispensing punishing stimuli in social interactions); and (5) Immature (dependent, high rates of initiation to teacher, inadequate social and study skills). Homogeneity and grouping will be determined by profile analyses which indicate high scores on one factor and low or moderate scores on the four remaining factors.

The Assessment Instrument

The WPBIC consists of fifty stimulus items that describe observable classroom behaviors. The fifty checklist items were drawn from teacher descriptions of classroom behavior problems. A random sample of thirty experienced teachers was drawn from the population of fourth, fifth, and sixth grade teachers in a local (Oregon) school district. The teachers were then asked to nominate those children in their classes who exhibited chronic behavior problems. Each teacher was then interviewed and asked to describe the child's behavior problem(s) and to give operational descriptions of the behaviors that concerned them. Observable descriptions of overt behavior were abstracted from each interview, yielding an item pool of three hundred items. Fifty of the most frequently mentioned behaviors from this sample were selected for inclusion in the checklist.

Items were assigned one of four score weights, from 1 to 4, indicating to what extent possession of a behavioral item handicaps the child's adjustment. Score weights were derived from a panel of behavioral scientists' ratings of the seriousness of the behavioral items in handicapping behavioral adjustment. Kuder-Richardson estimates of

the reliability of the WPBIC are .98 and .89 respectively (Walker, 1970). A test-retest reliability estimate with a one month interval yielded an r of .80 (Walker and Bull, 1970). The average item validity, as measured by correlations of individual items with total score, was .40. Contrasted groups validity indicates there was a statistically significant difference between the mean score of a group of deviant children and a matched group of normal children (N = 38). The biserial correlation between checklist scores and criterion scores, based upon three independent criteria of behavior disturbance, was .68. Consistent sex differences in checklist score were obtained across raters (teachers) and across grade levels.

The design of this study provided for the identification of factorially homogeneous groupings of pupils on five dimensions of classroom behavior. It also provided an opportunity for replication of results obtained with the normative sample upon another, larger sample of pupils in grades one, two, and three.

Results and Discussion

Comparisons Between Identification and Normative Samples

The WPBIC was standardized on a 534 pupil sample of children in grades four, five, and six. The identification sample consisted of 1067 children in grades one, two, and three. Table 1 contains the means and standard deviations for the two samples.

Insert Table 1

The difference of 3.02, in mean score, between the two samples is statistically significant beyond the .001 level.[1] The lower mean score of pupils in the identification indicates that as a group they were rated as less deviant by their teachers than pupils in the normative sample. Peterson (1961) has reported findings indicating the presence of non-linear developmental changes, as measured by behavior ratings, over the age range kindergarten through grade six. It is possible that the significant difference between the two samples reflects true developmental differences between pupils in grades one, two, and three and pupils in grades four, five, and six. However, acceptance of this hypothesis would mean that children exhibit significantly more deviant behavior as they progress through school. It would appear, at present that there is not enough data reported in the literature to provide conclusive support for this hypothesis.

Ross, Lacey, and Parton (1965) have suggested that when age-related changes on teacher checklists are found, they may be a function of systematic differences in teachers that are correlated with the grade level at which they are teaching. Support for this assumption is provided by Walker (1970) who found that teachers in grade six rated children in their classes as significantly less deviant than did teachers in grades four and five.

A more plausible explanation for the consistently lower scores of pupils in the identification sample relates to differences in sample seleccion procedures associated with the identification study and the original standardization study. In the standardization study, a random sample of classrooms at the fourth, fifth, and sixth grade levels was

drawn from the total number of elementary schools in the Eugene district. This procedure resulted in seven classrooms selected from each grade level. Teachers in the sample rated all pupils in their classrooms on the checklist. In the identification study, the school district required that signed permission slips be obtained from each child's parent prior to being rated on the checklist by his teacher. Thus all teachers in grades one, two, and three were included in the study. The return of permission slips and subsequent teacher ratings varied from zero to approximately seventy-five percent. Substantial feedback from teachers in the sample suggested that permission slips were not received from parents of the most deviant children in their classrooms. The fact that scores for these children were not included in the data analysis could explain the consistently lower scores of children in the identification sample.

Insert Table 2

Inspection of Table 2 reveals that the mean scores for pupils in grades four, five, and six are higher than the means for grades one, two, and three. The consistency of the effect across grades suggests that the mean scores for each grade level in the normative sample are more representative of the pupils' true behavioral status since they were based upon scores for all children enrolled in each classroom. If teacher reports that the more deviant children tended to be excluded from the sample are true; then checklist ratings on all pupils in each classroom in the identification sample would have probably resulted in higher mean scores for each grade level.

Intercorrelations Among the Factor Scales

The relationships that exist between the item clusters making up the five factors of the WPBIC are presented in the correlation matrices in Table 3.

Insert Table 3

Table 3 contains intercorrelations among the scales for both the identification and normative samples. The correlations indicate that with the exception of item clusters one and three, there is very little overlap among the five factor scales in both samples. This suggests that the WPBIC provides measures of separate dimensions within the same general behavior domain, e.g., behavioral disturbance.

The r of .67 between the acting-out and distractability factor scales in the normative sample and the equivalent r of .49 in the identification sample indicates that these two dimensions share the greatest amount of variance of any of the five factors within the checklist. The content of the items in each factor supports the assumption that the two scales measure common elements or dimensions of behavior. In addition, acting out or hyperactive children often manifest very high rates of non-attending and distractive behavior (Walker and Buckley, 1968; Patterson, Jones, Wright, and Whittier, 1965).

Intercorrelations among the five factor scales show a high degree of correspondence in the normative and identification samples. The coefficients, in Table 3, between scales four and five and between

scales one and three are identical in the two samples. The remaining intercorrelations for the identification sample closely parallel those for the normative sample in magnitude as well as in relative proportion to one another. Thus, relationships among the factor scales appear to remain stable across different samples of children.

Effects of Sex of Pupil and Grade Placement Upon Factor Scale Scores and Total Checklist Score

An analysis of variance for a 2 x 3 factorial design (Winer, 1962) was used to analyze the effects of grade and sex upon checklist score. Analyses of variance were computed for total score on the checklist and for each of the five, factor scales. Levels of each factor were male versus female and grade levels one, two, and three.

Insert Table 4

The F ratio of 29.61 in Table 4 indicates there was a statistically significant main effect for sex of pupil. The mean score for males across grade level was 5.97. The mean score for females was 3.63. Separate t tests indicated the sex difference was statistically significant within each of the three grade levels. There was no statistically significant effect for grade level. The interaction between grade level and sex of pupil was also not significant. The respective F ratios in Table 4 are 1.46 and .20. The significant sex difference in checklist score replicates an identical result obtained in the normative sample. This finding is also consistent with studies reported in the literature which indicate consistently more deviant scores for males than females

in ratings of child behavior (Quay, 1970). The finding of no statistically significant differences in checklist score between grade levels is consistent with results reported by Ross, Lacey, and Parton (1965). However, Peterson (1961) has reported the presence of nonlinear developmental changes, as measured by behavior ratings, over the age levels kindergarten through grade six. Ross, Lacey, and Parton point out that their results were based upon 31 teachers at each grade level while Peterson's results were obtained from an average of seven teachers at each grade level. Similarly, mean scores in this study were based upon an average of 47 teachers at each grade level. The issue of whether true developmental differences exist across grade levels is not clear at the present time. Ross, Lacey, and Parton argue that this issue must be resolved before behavior checklist data can be considered an unambiguous means of assessing developmental changes in the behavior of children.

Analyses of variance for each of the factor scales are presented in Tables 5 through 9.

Insert Tables 5 through 9

Inspection of the tables reveals that the statistically significant sex difference obtained for total checklist score held true for three of the five factors. Significant F ratios for sex of pupil were obtained for the acting out, distractability, and disturbed peer relations scales. The respective F ratios were 23.02, 67.19, and 4.55 respectively. The most powerful sex difference was associated with the acting out and distractability factors. Both these factors measure behaviors that

directly compete with academic performance. Thus, it appears these two clusters of behaviors clearly discriminate between males and females in the first three grades. There is considerable support for this hypothesis in the literature. Data from behavior rating scales, behavior checklists, and observation schedules indicate that these two response classes are powerful discriminators between male and female pupils. In addition, Buckley, Walker, Bridges, and Hendy (1970) operated a token economy classroom, over a four year period, for disturbed children. The most common reasons for referral were high rates of acting-out and/or distractable behavior. Of 65 children referred during this period, 59 were males and 6 were females.

Males were also rated as significantly more deviant than females on the cluster of items measuring disturbed peer relations. This factor provides a measure of the child's social relationship(s) with his peers. The behaviors making up the disturbed peer relations scale do not compete as directly with academic performance as do those comprising the acting out and distractability scales. However, possession of all or a substantial majority of the behaviors in the scale would severely handicap a child's educational as well as behavioral adjustment. If a true sex difference does exist on this cluster of behaviors, then it appears teachers are able to make valid discriminations on behaviors that directly compete with academic performance as well as those that handicap a child's educational adjustment in a less direct and more general way.

There was no statistically significant sex difference for the social withdrawal and immaturity factor scales. The F ratios for the main

effect of sex were .15 for social withdrawal and .06 for immaturity. These two factors measure behavior dimensions associated with restricted functioning, avoidance behavior, low rates of peer interactions, and peer relationships that would be classified as deficient or maladaptive instead of coercive or deviant. Thus, it would appear that males and females in grades one, two, and three share an approximately equal probability of being rated high on these two factor scales.

There was no significant main effect for grade level within any of the factor scales. In addition, there were no significant interaction effects for sex and grade within the five scales. Thus, the significant effect for sex of pupil and the absence of a significant effect for either grade level or interaction proved highly reliable across the factor scales, in this study.

Behavioral Incidence Data

The percentage of pupils receiving scores of one standard deviation above the mean (on each of the five scales) was analyzed for the identification sample. Separate analyses were conducted for each of the factors.

Insert Table 10

Table 10 contains the percentages of male, female, and total subjects scoring at or above the standard deviation for each factor scale using the means and sigmas for the original normative sample. An average of 5.90 percent of subjects scored at or above one sigma across the five scales. A z test was used to test the statistical significance of the

percentage difference between males and females scoring at or above one sigma within each factor scale. Table 10 reveals that the percentage differences for the acting-out, distractability, and disturbed peer relations scales were statistically significant. The social withdrawal and immaturity scale differences did not approach the levels required for significance. These data are consistent with the results of the analyses of variance of factor scores discussed earlier.

Insert Table 11

Table 11 contains the percentages of male, female, and total subjects scoring at or above one standard deviation for each factor scale using the means and sigmas for the identification sample. The percentages for male, female, and total subjects are larger due to the lower mean score(s) and smaller standard deviation(s) of the identification sample. However, the results in Table 11 replicate those in Table 10. An average of 11.34 percent of the total subjects scored at or above one sigma across the five factors in the identification sample. This compares with 5.90 percent using the normative sample means and sigmas. The statistically significant percentage differences for males and females are also identical in Table 11.

Summary and Conclusions

Results of this and other studies suggest that behavior checklist data provided by teacher ratings of child behavior provide a valuable and relatively inexpensive method of identifying homogeneous groupings of classroom behavior. However, the practice of relying upon a single

teacher rating to establish homogeneity and behavior class membership has not been clearly validated. It would appear a more intensive screening process using repeated observations of actual classroom behavior (with reliable observers) would be necessary to reliably determine homogeneity. Further, the use of checklist data to measure developmental changes and to evaluate the effects of intervention does not appear to be justified by research data presented in the literature.

Results of this study indicate that teacher ratings of various classes of behavior reflect sex differences that have been validated in other studies. The data appear to be internally consistent and replicate many of the results obtained with the normative sample.

References

Becker, W. C. The relationship of factors in parental ratings of self and each other to the behavior of kindergarten children as rated by mothers, fathers, and teachers. Journal of Consulting Psychology, 1960, 24, 507-527.

Becker, W. C. and Krug, R. S. A circumplex model for social behavior in children. Child Development, 1964, 35, 371-396.

Beilin, H. Teachers' and clinicians' attitudes toward the behavior problems of children--A reappraisal. Child Development, 1959, 30, 9-25.

Boneau, C. A. The effects of violating of assumptions underlying the t-test. Psychological Bulletin, 1960, 57, 49-64.

Bower, E. M. Early identification of emotionally handicapped children. Springfield, Illinois: Charles C. Thomas, 1960.

Buckley, N. K., Walker, H. M., Bridges, D., and Hendy, M. Construction and evaluation of a treatment model for modifying deviant behavior in children. Section #3. Final Report for U.S.O.E. Project OEG 4-6-061308-0571. Assessment and Treatment of Deviant Behavior in Children, Bureau of Education for the Handicapped, Division of Research.

Clark, E. J. Teacher reactions toward objectionable pupil behavior. Elementary School Journal, 1951, 51, 446-449.

Cobb, J. A. Survival skills and first grade academic achievement. Report #1, Early Childhood Education Center, University of Oregon, 1970.

Cowen, E. L., Zax, M., Izzo, L. D., and Trost, M. Prevention of emotional disorders in the school setting. Journal of Consulting Psychology, 1966, 30, 381-387.

Cromwell, R. Behavior classification of emotionally disturbed children. Paper presented at Council for Exceptional Children Convention, Portland, Oregon, April, 1965.

Downie, N. M. and Heath, R. W. Basic Statistical Methods. New York: Harper and Row, 1970, 3rd edition.

Dreger, R. Behavioral classification project. Journal of Consulting Psychology, 1964, 28, 1-13.

Dunn, L. M. Special education for the mildly retarded--is much of it justifiable? Exceptional Children, 1968, 35, 5-22.

Edwards, A. L. Statistical Methods for the Behavioral Sciences. New York: Holt, Rinehart, and Winston, 1967, 2nd edition.

Fitzsimmons, S., Cheever, J., Leonard, E., and Macunovich, D. School failures: Now and tomorrow. Developmental Psychology, 1969, 1, 134-146.

Haring, N. and Ridgway, R. W. Early identification of children with learning disabilities. Exceptional Children, 1967, 33, 387-395.

Hunter, E. C. Changes in teachers' attitudes toward children's behavior over the last thirty years. Mental Hygiene, 1957, 41, 3-11.

Kennedy, D. A study of the academic achievement of 4th grade boys relative to four measures of emotional adjustment. Unpublished thesis, University of Oregon, 1965.

Kulik, J. A., Stein, K., and Sarbin, T. Dimensions and patterns of adolescent antisocial behavior. Journal of Consulting Psychology, 1968, 32, 375-383.

Haes, W. R. The identification of emotionally disturbed children. Exceptional Children, 1966, 32, 607-613.

O'Neal, P. and Robbins, L. N. The relation of childhood behavior problems to adult psychiatric status: A 30-year follow-up of 150 subjects. American Journal of Psychiatry, 1958, 14, 961-969.

Packard, R. G. The control of classroom attention: A group contingency for complex behavior. Journal of Applied Behavior Analysis, 1970, 3, 13-28.

Patterson, G. R. An empirical approach to the classification of disturbed children. Journal of Clinical Psychology, 1964, 20, 326-337.

Peck, L. Teachers' reports of the problems of unadjusted school children. Journal of Educational Psychology, 1935, 26, 123-138.

Peterson, D. R. Behavior problems of middle childhood. Journal of Consulting Psychology, 1961, 25, 205-209.

Peterson, D. R. Scope and generality of verbally defined personality factors. Psychological Review, 1965, 72, 48-59.

Phillips, B. Problem behavior in the elementary school. Child Development, 1968, 39, 895-903.

Quay, H. C. Dimensions of personality in delinquent boys as inferred from the factor analysis of case history data. Child Development, 1964, 35, 479-484.

Quay, H. C. The education of behaviorally disordered children in the public school setting. Final Report, Project No. 482207, Grant No. OEG 3-6-62-63-1559 (032) Bureau of the Handicapped, Division of Research, U. S. Office of Education.

Quay, H.C. The facets of exceptionality: A conceptual framework for assessment, grouping, and instruction. Exceptional Children, 1968, 32, 297-301.

Quay, H. C., Morse, W. C., and Cutler, R. L. Personality patterns of pupils in special classes for the emotionally disturbed.

Quay, H. C. and Quay, L. C. Behavior problems in early adolescence. Child Development, 1965, 36, 215-220.

Robins, L. N. Deviant children grown up: A sociological and psychiatric study of sociopathic personality. Baltimore: Williams and Wilkins, 1966.

Ross, A. O., Lacey, H., and Parton, D. The development of a behavior checklist for boys. Child Development, 1965, 36, 1013-1027.

Sarason, E. H., Davidson, K. S., Lighthall, F. K., Waite, R. R., and Ruebush, B. K. Anxiety in Elementary School Children. New York: John Wiley and Sons, 1960.

Schrupp, M. H. and Gjerde, C. N. Teacher growth in attitudes toward behavior problems of children. Journal of Educational Psychology, 1953, 44, 203-214.

Sines, J. O., Pauker, J. D., Sines, L. K., and Owen, D. R. The identification of clinically relevant dimensions of children's behavior. Journal of Consulting Psychology, In Press.

Spivack, G. and Swift, M. S. The Devereux elementary school behavior rating scales: A study of the nature and organization of achievement related disturbed classroom behavior. Journal of Special Education, 1966, 1, 71-90.

Stennett, R. G. Emotional handicap in the elementary years: Phase or disease. American Journal of Orthopsychiatry, 1965, 36, 444-449.

Stouffer, G. A. Behavior problems of children as viewed by teachers and mental hygienists: A study of present attitudes as compared with those reported by E. K. Wickman. Mental Hygiene, 1952, 36, 271-285.

Swift, M. S. and Spivack, G. The assessment of achievement related classroom behavior. Journal of Special Education, 1968, 2, 137-154.

Trippe, M. Conceptual problems in research on educational provisions for disturbed children. Exceptional Children, 1961

Ullmann, C. A. Identification of maladjusted school children, Public Health Monograph No. 7, Washington D. C., 1952. Federal Security Agency.

Walker, H. M. Empirical assessment of deviant behavior in children. Psychology in the Schools, 1969, 6, 93-97.

Walker, H. M. Manual for the Walker problem behavior identification checklist. Section #2, Final Report for U.S.O.E. Project OEG 4-6-061308-0571 Assessment and Treatment of Deviant Behavior in Children. Bureau of Education for the Handicapped, Division of Research.

Walker, H. M. and Buckley, N. K. Investigation of some classroom control parameters as a function of teacher dispensed, social reinforcers. Submitted for publication to: Journal of Applied Behavior Analysis, 1970.

Walker, H. M., Mattson, R. H., and Buckley, N. K. The functional analysis of behavior within an experimental class setting. Chapter in W. C. Becker (Ed.) An Empirical Basis for Change in Education, Science Research Associates, Inc. (In Press).

Watson, G. A critical note on two attitude scales. Mental Hygiene, 1933, 17, 59-64.

Westman, J. C., Rice, D. L., and Bernamn, E. Relationships between nursery school behavior and later school adjustment. American Journal of Orthopsychiatry, In Press.

Wickman, E. K. Children's Behavior and Teacher's Attitudes. New York: The Commonwealth Fund, 1928.

Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill Company, 1962.

Zax, M. and Cowen, E. L. Early identification and prevention of emotional disturbance in a public school. In E. L. Cowen, E. A. Gardner, and M. Zax (Eds.) Emergent Approaches to Mental Health Problems. New York: Appleton-Century-Crofts, 1967.

Zax, M., Cowen, E. L., Izzo, L. D., Madonia, A. J., Merenda, J. and Trost, M. A. A teacher aide program for preventing emotional disturbance in young school children. Mental Hygiene, 1966, 50, 406-415.

Zax, M., Cowen, E. L., Izzo, L. D., and Trost, M. A. Identifying emotional disturbance in the school setting. American Journal of Orthopsychiatry, 1964, 34, 447-454.

Zax, M., Cowen, E. L., Rappaport, J., Beach, D., Laird, J. Follow-up study of children identified early as emotionally disturbed. Journal of Consulting Psychology, 1968, 32, 369-373.

Footnotes

1. A test applied to the variances of the identification sample and the normative sample indicated the assumption of homogeneity of variance underlying the t test could not be met. Boneau (1960) argues that if two samples have unequal sizes and unequal variances and their respective distributions are skewed (as in this case), the resulting t ratios will also tend to be skewed and will lead to biased results. However, using samples of larger size tends to remove this skew (Downie and Heath, 1970). Further, Edwards (1967) notes that if the t test is applied to independent random samples of size 25 or more, the t test is relatively unaffected by rather severe violations of the assumptions of homogeneity of variance and normality of the distributions in the population. Considering the robustness of the t test and the size of the two independent samples in this study, it was decided to use the t test to determine statistical significance of the mean difference.

Table 1

Means and Standard Deviations for the Identification and Normative Samples

Identification Sample (N=1067)		Normative Sample (N=534)			
$\bar{X}$	S.D.	$\bar{X}$	S.D.	D.	Critical Ratio
4.74	6.66	7.76	10.53	(3.02)	6.16***

*Significant at .05
**Significant at .01
***Significant at .001

Table 2

Mean Scores for Pupils in the
Identification and Normative Samples

	Identification Sample (N=1067)			Normative Sample (N=534)		
Grade	1	2	3	4	5	6
$\bar{X}$	4.97	5.00	4.20	9.48	8.72	5.04

Table 3

Intercorrelations of the Five WPBIC Factor Scales in the Identification Sample and the Normative Sample*

	Acting-Out	Social Withdrawal	Distracta-bility	Disturbed Peer Relations	Immaturity
Acting-Out	--	.02 (.09)	.67 (.49)	.48 (.37)	.39 (.28)
Social Withdrawal			.12 (.12)	.18 (.28)	.23 (.32)
Distracta-bility				.48 (.31)	.44 (.28)
Disturbed Peer Relations					.34 (.34)
Immaturity					--

*Intercorrelations within parentheses are for the identification sample. Unenclosed coefficients are for the normative sample.

Table 4

Analysis of Variance for a 2 x 3
Factorial Design: Total Checklist Score

Source	SS	DF	MS	F
Total	47,274	1,066	--	--
Sex of pupil (A)	1,279	1	1,279	29.61***
Grade level (B)	126	2	63	1.46
A x B	17	2	8	.20
Error	45,852	1,061	43	

*Significant at .05
**Significant at .01
***Significant at .001

Table 5

Analysis of Variance for a 2 x 3
Factorial Design: Acting-Out Scale

Source	SS	DF	MS	F
Total	9,577	1,066	--	--
Sex of Pupil (A)	202	1	202	23.02***
Grade level (B)	29	2	14	1.66
A x B	14	2	7	.80
Error	9,332	1,061	8	

*Significant at .05
**Significant at .01
***Significant at .001

Table 6

Analysis of Variance for a 2 x 3
Factorial Design: Social Withdrawal Scale

Source	SS	DF	MS	F
Total	2,910.05	1,066	--	--
Sex of pupil (A)	.40	1	.40	.15
Grade level (B)	5.72	2	2.86	1.05
A x B	2.51	2	1.25	.46
Error	2,901.42	1,061	2.73	

*Significant at .05
**Significant at .01
***Significant at .001

Table 7

Analysis of Variance for a 2 x 3
Factorial Design· Distractability Scale

Source	SS	DF	MS	F
Total	6,398	1,066	--	--
Sex of pupil (A)	380	1	380	67.19***
Grade level (B)	11	2	5	1.03
A x B	6	2	3	.61
Error	6,001	1,061	5	

*Significant at .05
**Significant at .01
***Significant at .001

Table 8

Analysis of Variance for a 2 x 3
Factorial Design: Disturbed Peer Relations Scale

Source	SS	DF	MS	F
Total	2,129	1,066	--	--
Sex of pupil (A)	9	1	9	4.55*
Grade level (B)	3	2	1	.86
A x B	1	2	.50	.25
Error	2,116	1,061	1.99	

*Significant at .05
**Significant at .01
***Significant at .001

Table 9

Analysis of Variance for a 2 x 3
Factorial Design: Immaturity Scale

Source	SS	DF	MS	F
Total	1,957.08	1,066	--	--
Sex of pupil (A)	.11	1	.11	.06
Grade level (B)	6.97	2	3.49	1.90
A x B	.53	2	.26	.14
Error	1,949.47	1,061	1.84	

*Significant at .05
**Significant at .01
***Significant at .001

Table 19

Percentage of Subjects Receiving Scores at or Above One Standard Deviation Above the Mean(s) of the Normative Sample on Each Factor Scale

	Acting-Out	Social Withdrawal	Distracta-ability	Disturbed Peer Relations	Immaturity
% of total sample scoring at or above 1 sigma	5.25	3.00	6.28	5.43	9.56
% of male SS scoring at or above 1 sigma	4.22	1.50	5.06	3.56	4.97
% of female SS scoring at or above 1 sigma	1.03	1.50	1.22	1.87	4.59
% difference	4.19***	--	3.84***	1.69*	.38

*Significant at .05
**Significant at .01
***Significant at .001

Table 11

Percentage of Subjects Receiving Scores at or Above One Standard Deviation Above the Mean(s) of the Identification Sample on Each Factor Scale

	Acting-Out	Social Withdrawal	Distracta-ability	Disturbed Peer Relations	Immaturity
% of total sample scoring at or above 1 sigma	10.31	8.34	14.81	9.75	13.49
% of male SS scoring at or above 1 sigma	7.31	4.22	11.53	6.09	7.40
% of female SS scoring at or above 1 sigma	3.00	4.12	3.28	3.75	6.09
% difference	4.31***	.10	8.25***	2.25*	1.31

*Significant at .05 level
**Significant at .01 level
***Significant at .001 level

Appendix I

MANUAL, RATING INSTRUCTIONS, AND CODING CRITERIA FOR THE OBSERVATION SCHEDULE FOR ACTING-OUT BEHAVIOR

Oregon Center for Research and Demonstration
in the Early Education of Handicapped Children
Department of Special Education, College of Education
University of Oregon, Eugene, Oregon 97403

I. Coding Instructions for the Observer

The following descriptions of observation methods and data recording procedures are designed to log the acting-out child's behavior in the classroom setting, and to record consequences supplied to that behavior. The form provides a record of behavior, measures rate, and notes consequent social responses to the child's behavior from the environment. The form also produces data on the child's performance in the classroom, and relates this data to consequating events supplied by the teacher and peers. The purpose of the observation form is three-fold:

1. To identify the acting-out child in the classroom setting.
2. To record the social consequences of the acting-out child's behavior.
3. To monitor the behavior of his peers during observation sessions for purposes of comparative analysis.

The observation form is divided into three two-minute sections representing six minutes of observation. The "target" child may be compared to his peers, for monitoring purposes, by alternating one sheet on a "normal" peer chosen at random by the observer. (This is the convention used by our staff. However, it is permissible to use other ratios of monitoring; i.e., 2 sheets or 12 minutes on the "target" child, and 1 sheet or 6 minutes of data on a selected peer.) Each observation form supplies the following information:

1. Behavior of the subject.
2. Rate of behavior.
3. Whether the behavior is appropriate or inappropriate to the situation.
4. Consequent response to the behavior.
5. Agent supplying consequent response.
6. Description of the classroom situation.

Each two minute section has eight horizontal lines of squares. Each line represents 15 seconds of observation. Each square in the line is a behavior class, with some squares divided in half. Coding in the top half signifies an inappropriate behavior, while coding in the bottom half indicates an appropriate behavior. The general procedure is to observe the subject during the interval, and code the behavior and consequent responses in the process. At the end of the 15-second interval, the observer drops down to the next line while continuing to observe and code the interactions. The interval is marked by a timing device set into a clipboard which buzzes every 15 seconds. When this occurs, it marks the start of a new interval, and the observer starts the next line. (The staff tried a 15 second interval in which 10 seconds were used for observation only, and 5 seconds were for recording interactions; however, the 15 second interval of continuous observation and recording produced higher reliability among observers.)

The squares in each line correspond to the coded behaviors listed above the individual square. The observer indicates a behavior has occurred by placing the consequent response to that behavior and response agent in the square describing the subject's behavior, while noting whether the behavior was appropriate or inappropriate. Appropriate behavior is defined by the environment and the rules operating in the classroom. The observer should find out from the teacher when, and under what conditions, behaviors, such as talking to a peer, are appropriate. Other behaviors may be more obvious; i.e., during Math, the subject may initiate a vocalization (VO) to the teacher (T), asking her whether his answer is correct; the teacher responds by praising the subject for having the right answer (P). The correct coding procedure is illustrated below:

WK	NO	NA	MY	VO	PH	MO	IS
PT				+TPT			+PT

Coding in the preceeding example is marked in the following sequence: the subject's behavior, the agent recipient (VO and PH are the only categories in which the subject addresses peer or teacher), the response to the subject's behavior, and the agent performing the response to the behavior.

The behavior, consequent response, and agents are listed with their codes above each observation form. To avoid unnecessary confusion at the beginning of training, it is suggested the prospective observer familiarize himself with these codes and definitions provided on the following pages. It is not necessary to make more than one notation of the same behavior within a 15 second interval. The observation form is designed to point out the subject's behavior and social consequences. Sequencing is not the important variable.

After mastering the behavior, consequence, and agent codes, the potential observer may wish to gauge his progress in mastering the method or process of observation: a criterion test is included at the end of the manual for this purpose. Behavioral interactions that are normally encountered in the classroom are used for the criterion test. The reader may want to test his ability in coding these interactions on the first observation form. The second form has been coded by our staff on the same interactions so that the reader may compare his coding against ours. This is an easy check that will test understanding of the manual and help to clarify confusions about the observation codes.

II. Observation Code Criteria

Classroom

Group: To be marked whenever the class is participating as a whole i.e., class lecture, reading, and listening. This does not apply to class projects where individual work is required.

Individual: To be marked whenever the subject is involved in separate study; i.e., reading workbooks, math problems, and art. The teacher is helping individual children and the subject is expected to do his own work.

Transition: To be marked whenever the entire class changes from one activity to another; i.e., clearing desks for recess, stopping one lesson and starting another, forming groups for reading. Movement is usually characterized in transition. When movement occurs during transition, it should be double coded as normative (NO) as well as movement (MO).

Description Of Codes:

Classroom Behaviors:

WK (Individual Work): Appropriate - Inappropriate
Appropriate work is coded in the lower half of the square whenever the subject is engaging in the prescribed individual work: The class is working in its Math workbooks and the subject is involved in the same activity - he has the workbook in front of him and is attending to the problems. Inappropriate work is coded in the upper half of the square whenever the subject is engaging in activity other than the prescribed work indicated by the teacher. The teacher has told the class to read the story on page 25 of their readers. The subject, instead, continues his drawing from the last period. It is important to note that the observer should not code non-attending (NA) while coding inappropriate work, such as cheating, flipping through pages unnecessarily, or sharpening a pencil for a long length of time, unless the subject is not attending to this inappropriate work.

NO (Group Activity): To be coded whenever subject engages in behavior characterized by group activity or transitional phases: i.e., moving chairs to form reading group, raising hand to answer or ask a question, putting away math book and getting out reading book, lining up for recess or fire drill, listening to teacher's instructions and group discussions. It is important to note that a child may be doing individual work within special study groups, unless the members are working together, this should be coded (WK).

NA (Nonattending): To be coded whenever the subject is not attending to his work or class lesson. This may occur during WK or NO, when the subject should be attending. This involves looking up from his work on his desk, looking out the window during a class lecture, or resting his head on the desk while he should be attending. This behavior is always classified as inappropriate.

NY (Noise - non verbal): To be coded whenever the subject engages in loud, disruptive noises; i.e., banging book on desk, kicking desk, mumbling, and incoherent utterances. This category is always coded as inappropriate and usually accompanies movement (MO).

VO (Vocalization): Appropriate - Inappropriate
To be coded whenever the subject is engaged in coherent vocalizations that are not considered noise; i.e., singing, talking to one's self or to others.
Appropriate vocalizations are coded in lower half of the square and include: talking to teacher or peer with permission.
Inappropriate vocalizations are coded in the upper half of the square and include: talking to peer about topics other than the lesson, talking without permission, talking to disrupt and annoy others in the class.
While coding VO square, the observer must record:

1. The agent addressed by the subject.
2. The agent's response to the subject.
3. The agent who responded to the subject.

In most cases, the agent is the same in items one and three: however, the subject may direct a statement to a peer but the teacher actually responds by disapproval, in this case, the code will look like this PDT under inappropriate VO. Other cases may show double responses to the same VO, as in the example above, the peer attends to the subject and the teacher disapproves, the coding will look like this APDT.
(Frequently, a child may mouth words while reading; this is not considered vocalization.)

PH (Physical Contact + or -): Appropriate - Inappropriate
To be coded whenever subject engages in physical contact with others. Contact may be regarded as either positive (+): i.e., placing an arm around peer, or negative (-); i.e., striking peer or teacher.
Inappropriate contact is coded in the upper half of the square and indicates annoying or disruptive behavior or is inappropriate at the time: The subject touches every peer he passes returning to his desk after sharpening his pencil; or he taps peer sitting in front of him for attention.
Appropriate contact is coded in the lower half of the square whenever subject touches others in a situation permitting contact; i.e., games.
Aggression, actual or attempted, is coded as negative (-) and inappropriate: Subject strikes peer or attempts to strike peer but is stopped by the teacher.

MO (Movement): Appropriate - Inappropriate
To be coded whenever subject is moving in his chair; i.e., squirming, turning around, raising hand, or following motor instructions from teacher. Movements are considered appropriate or inappropriate: Getting out of seat or not touching seat and standing at desk are coded by placing an X in the square under Movement.

IS (Vocal Initiation to Subject):
To be coded whenever an agent, either the teacher (T) or a peer (P), addresses the subject or replies to the subject. Anways code

under IS in this manner:

1. code the agent involved with the initiation.
2. code the subject's response to the agent's initiation.

Appropriate initiations are coded in lower half of IS square and entail statements or dialogues between teacher and subject and talking between peer and subject approved by the teacher. Inappropriate initiations are characterized by a peer disrupting the subject from his work or conversations between peer and subject without teacher permission.

Responses:

A (Attention): To be coded whenever agent attends to specified behavior. This is considered a neutral response, void of approval or disapproval: The teacher looks at or listens to the subject.

P (Praise): To be coded whenever response agent displays approval of subject's behavior. This may be through a verbal response or a gesture: i.e., "That's nice," head nod, smile.

D (Disapproval): To be coded whenever peer or teacher indicates disgust or disapproval of subject's behavior. Responses may be verbal or gestural: i.e., "Don't do that!", head shake, frown.

O (Ignore): To be coded whenever a behavior occurs by the subject and there is no response from the teacher or peers. Under IS, an agent may initiate to the subject and be ignored.

C (Compliance): The subject responds to teacher or peer initiated command.

NC (Noncompliance): The subject does not comply with teacher or peer initiated command.

PH (Physical + or -): To be coded whenever response agent responds to subject either by positive contact: hugging, patting; or by negative contact: hitting.

Response Agents:

Each response must specify which agent is involved. The agents are coded as follows:

T - teacher
P - peer
O - observer

The observer is encouraged not to interact with members of the class, and to ignore questions by the students. O should rarely be used as an agent since he is not an integral member of the class.

III. Coding Situations for Criterion Test

Situation	Correct	Incorrect
1. The teacher assigned a math sheet for the students to do individually. S. is at his desk doing his work. He looks over at his neighbor to copy an answer. The teacher doesn't see the student cheat. He continues working.	____	____
2. S is working on an individual activity at his desk. The observer in the back of the room drops his clipboard. S is startled by the noise and turns around to see where it came from. He then returns to his work. No one noticed him look at the observer.	____	____
3. S is receiving help from the teacher on his math. The teacher asks him a question which he answers. The teacher says, "Very good!: Another student approaches the teacher with a question. S waits watching the teacher and other student talk. The teacher then turns back to S to help him with his work.	____	____
4. S is working at his desk on an art project. He looks up when he hears someone talking, and then returns to his work. To see his work better, he stands up and works. No one talks to him or gives him attention.	____	____
5. S is asked by the teacher to pass out work to be taken home. S gets the papers and begins passing them out. He stops and reads each paper before giving them to the student. No one talks to him or gives him attention.	____	____
6. The teacher has given an assignment for the students to write a story at their desks. S breaks his pencil accidently and goes to sharpen it. (Permission is not necessary for pencil sharpening.) He returns to his desk and begins writing with his head laid on his left arm as he writes.	____	____
7. The teacher is reading a story to the class. S slides his chair, without leaving his seat, resulting in a scraping sound. He begins burping loudly over and over. The teacher looks at him and shakes her head.	____	____
8. The class is having a discussion about birds. S sits near the window. The teacher asks a question. S raises his hand and yells, "I know it." The teacher		

ignores him and calls on someone else. S leaves his hand up and gradually lets it fall behind him until he grabs the window shade. He pulls the shade so it goes up with a loud bang. The teacher ignores him while a peer looks over and laughs. ____ ____

9. During math period, the teacher assigns a page of individual work. S works two problems; then gets up from his seat and goes to ask the teacher about a problem. She tells him the answer and he thanks her. On his way back to his desk, he stops at another student's desk and asks about how his work is going. The student tells him he only has ten problems to go. S shows him his work; then returns to his desk. ____ ____

10. S is sitting next to the wall, in his chair, beyond reach of his desk. The teacher assigns him a book to read at his desk. He leans back in his chair until he bangs the wall. He bangs repeatedly until the teacher says, "Stop that and do your work." He replies, "I don't have to," but does stop the banging. He gets up, moves his chair back to his desk and sits down, beginning to read his book. ____ ____

11. S is sitting at his desk during a class discussion of what causes accidents. He listens to the teacher say something about airplanes. He then puts his hand in the air as if it were an airplane crashing into his desk, and makes the sound effects of the crash. The boy next to him notices, and begins imitating him. The teacher ignores them. ____ ____

12. During a spelling test, S looks at the teacher until she gives a word, then writes it on his test paper. He then reaches down to scratch his leg, and looks back at the teacher to wait for the next word. ____ ____

13. The teacher indicates to the class that they are to read Chapter I of their text. S begins to read Chapter I; then begins reading captions under the pictures throughout the book. No one seems to notice his activity. ____ ____

14. Two minutes ago, S was conscientiously involved in doing his math assignment. He then began sharpening his pencil. Now he is still involved in sharpening his pencil, while he is discussing his family with a peer. The peer asks a question regarding S's family. ____ ____

15. S is writing a paper on birds: the assigned task. He is seated sideways on his chair when a peer returns to the classroom. S looks up when the door opens; then begins to write again while shifting his body to the "normal" desk-sitting position. A peer asks S the answer to number 8. S ignores the peer and continues to work. ____ ____

16. S is raising his hand to answer a question put to the class. The teacher does not call on S. The selected peer answers, but is incorrect. S then answers the question without teacher permission. The teacher tells S to raise his hand if he has something to say. S makes no response.

___ ___

17. The class is involved in a spelling test. The teacher pronounces a word; the students then write the word and wait for the next word. While taking the test, S is seated sideways in his chair. After writing one of the words, he begins tapping his pencil on the desk. A peer begins to smile and imitate S's tapping behavior. The teacher frowns and S discontinues the activity.

___ ___

18. While sitting in a group during show and tell, S begins pulling his shirt over his head and making a deep humming sound. Three peers laugh at his shirt pulling and noise. The teacher tells S to return to his seat because he doesn't know the proper behavior. S returns to his seat.

___ ___

19. The class is watching a humorous film. During a funny scene, S begins to laugh as do other class members. S begins to jab a peer in the ribs. The peer says "Don't" and jabs S back. S ignores the peer and continues watching the film.

___ ___

20. During art, S is involved in reading a book on history. The teacher tells the class members to clean their desks and line up at the door for recess. S clears his desk and walks directly to the line, where he begins talking with a friend. The peer is listening. S then walks over to an "interest" center and begins to examine the supplies there. No one notices him.

___ ___

21. After "free-time" the teacher instructs the students to go to their seats and work on their Math assignment. S begins to work on the assigned task; but he is not seated at his own desk.

___ ___

22. S is still not seated at his own desk (see #21). S is working on his Math, while resting his head on his desk. The teacher notices that S is not at his own desk, and asks S to return to his own seat. S walks directly to his desk.

___ ___

23. The class has been instructed to work on its new science assignment. S and a peer have been talking about baseball. S is seated with his

chair tilted back on two legs and intermittently whistles. The peer smiles whenever S whistles. The teacher tells S to stop whistling. S complies. ____ ____

24. S is singing the selected song during Music. After the song, he listens to the explanation of a foreign song; then, with the class, begins to sing the next song. ____ ____

Total Score

# Correct	# Incorrect
____	____

IV. OBSERVATION FORM AND ANSWERS TO CRITERION CODE TEST

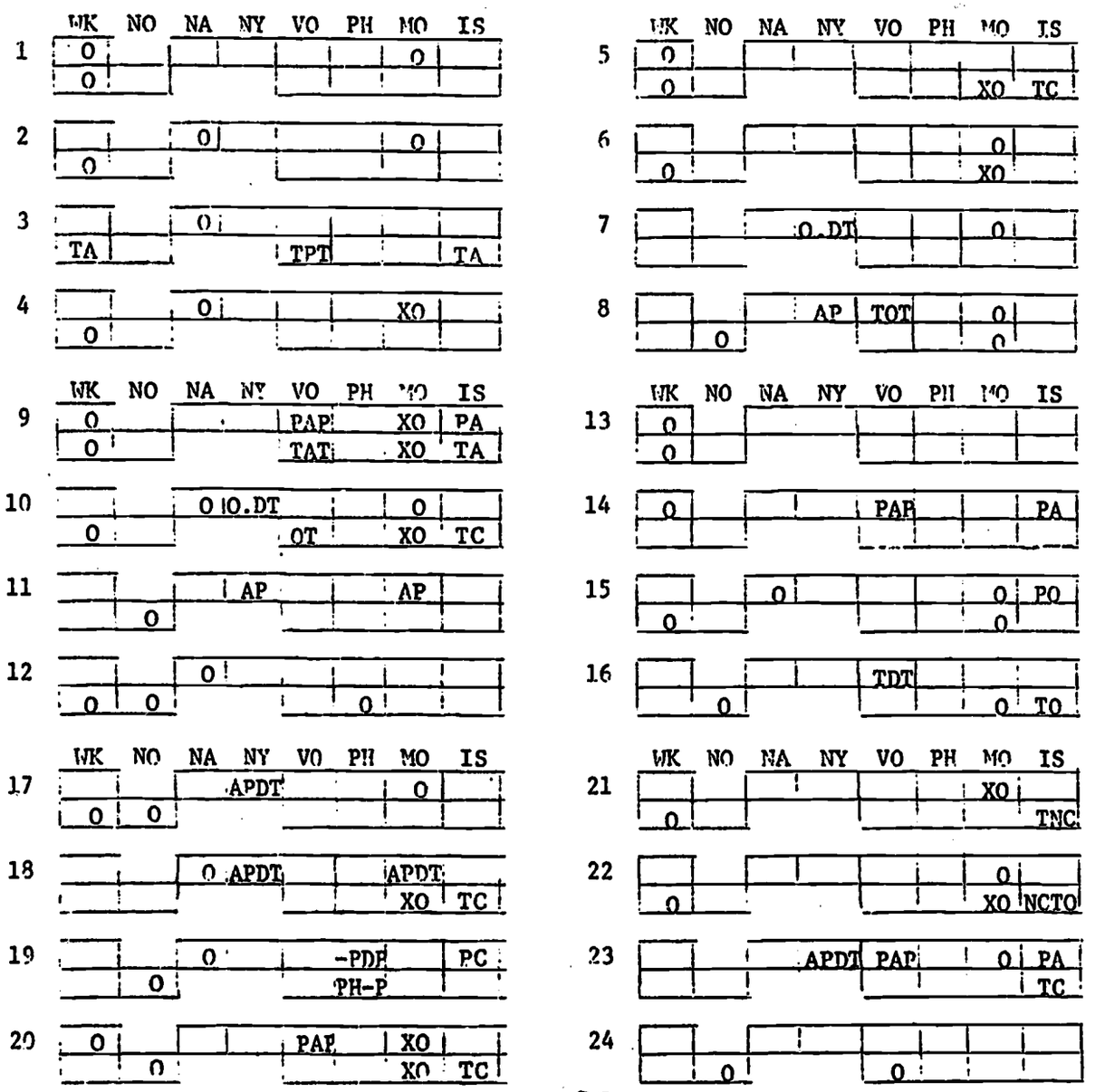
Student_____ Observer_____ Date_____

Group_____ Individual_____ Transition_____ Subject_____ Sheet No._____ Time_____

Behaviors:
WK--Individual work
NO--Group activity
NA--Nonattending
NY--Noisy (Not vocal)
VO--Vocalization
PH--Physical contact (+ or -)
MO--Movement X--Out of seat
IS--Initiation to subject

Responses:
A--Attention
P--Praise
D--Disapproval
O--Ignore
C--Compliance
NC--Noncompliance
PH--Physical (+ or -)

Agents:
T--Teacher
P--Peer
O--Observer



IV. OBSERVATION FORM AND ANSWERS TO CRITERION CODE TEST

Student_____ Observer_____ Date_____

Group_____ Individual_____ Transition_____ Subject_____ Sheet No._____ Time_____

Behaviors:

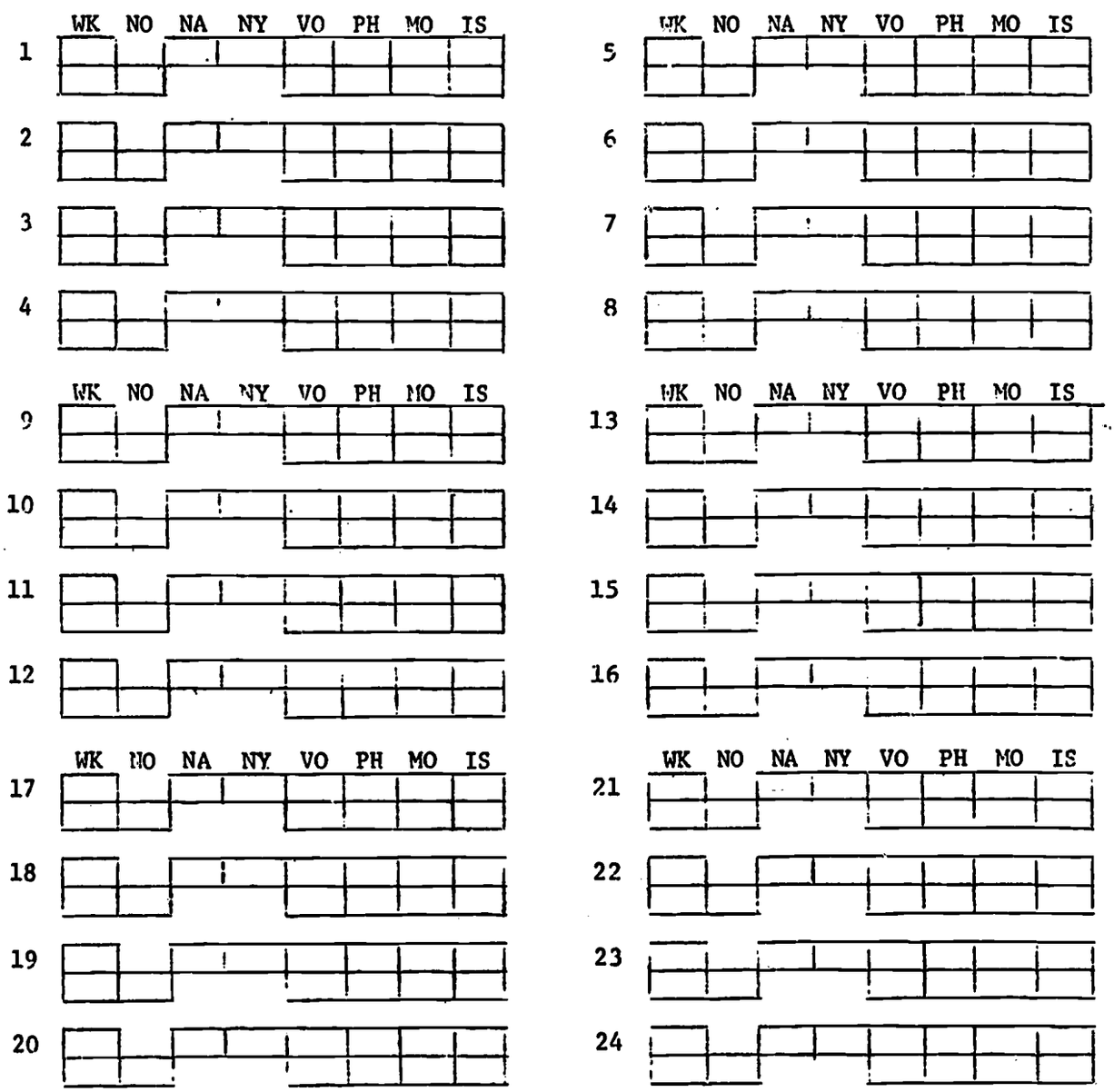
WK--Individual work
NO--Group activity
NA--Nonattending
NY--Noisy (not vocal)
VO--Vocalization
PH--Physical contact (+ or -)
MO--Movement X--Out of seat
IS--Initiation to subject

Responses:

A--Attention
P--Praise
D--Disapproval
O--Ignore
C--Compliance
NC--Noncompliance
PH--Physical (+ or -)

Agents:

T--Teacher
P--Peer
O--Observer



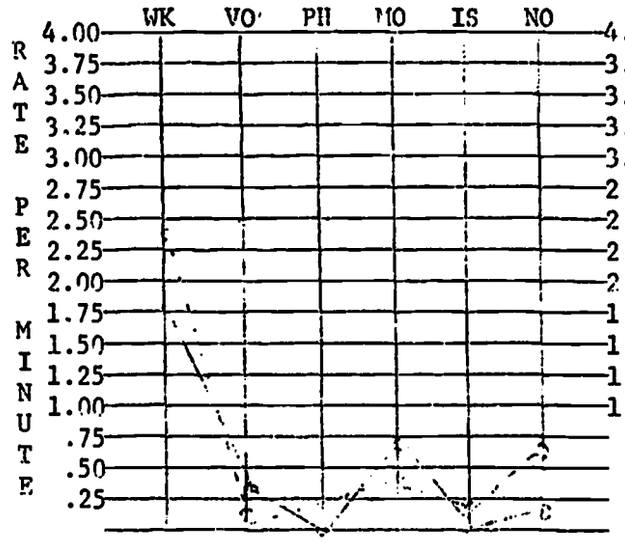
V. Normative Data

Since the observation schedule was developed specifically for measuring acting-out behavior in the classroom; it should discriminate between those children whose behavior can be characterized as acting-out, and those whose behavior is not so characterized. Using a combination of teacher nomination, checklist scores, and observation data, six acting-out children were selected as experimental subjects. Three observers were trained to a criterion of .90 inter-subject reliability on the observation schedule. Observations were taken on the acting-out subject, and a randomly selected peer, until the observers had recorded one hour of observation data on the experimental subject and one hour on his peers. Two observers recorded simultaneously: one recorded the peer's behavior, and the other recorded the acting-out child's behavior. Ten sheets were recorded on the student and one sheet on each of ten peers. Each sheet covered six minutes totaling twenty-four fifteen-second intervals. The first peer in the front row was observed first. The observer proceeded down the row; then began the next row, until ten sheets were completed. The following charts compare the rates of the student and his peer's average rate for appropriate and inappropriate behavior.

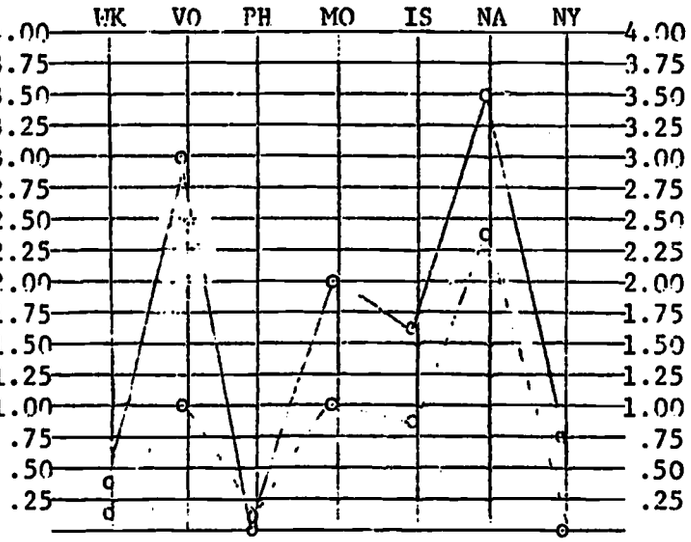
COMPARATIVE OBSERVATION DATA FOR
ACTING-OUT SUBJECTS AND THEIR AVERACE PEERS

S Student I

Appropriate



Inappropriate



Categories

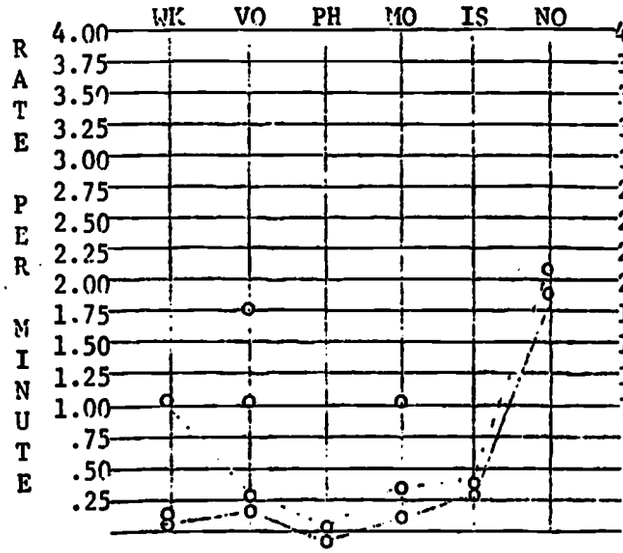
Experimental Subject

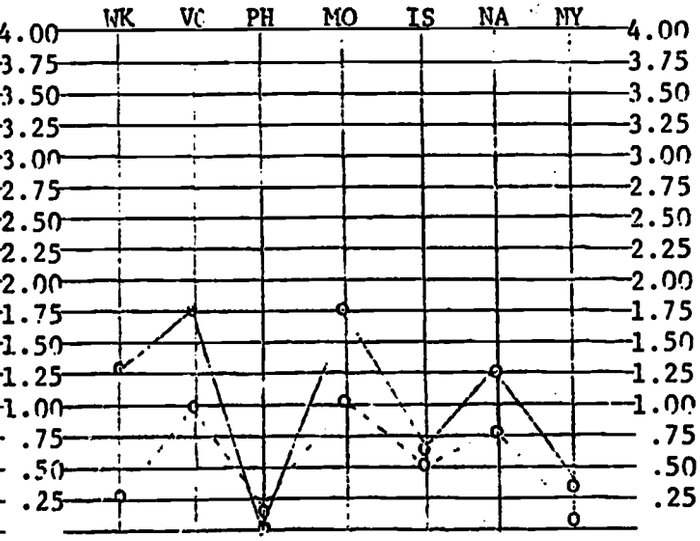
Average Peer

Categories

Student II

Appropriate

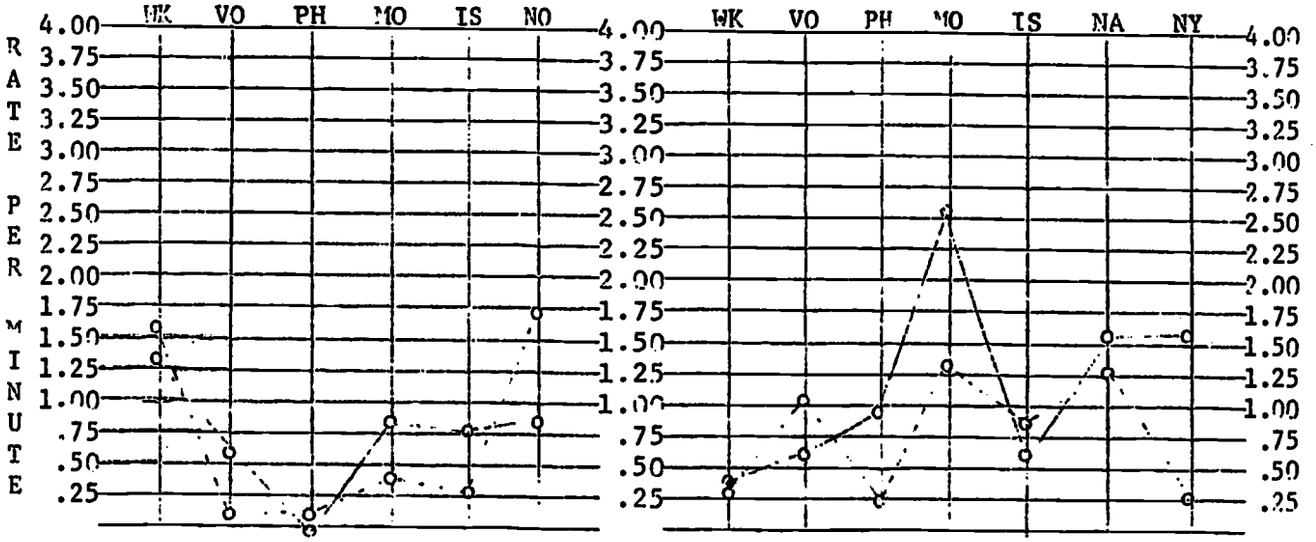




Student III

Appropriate

Inappropriate



Categories

Student

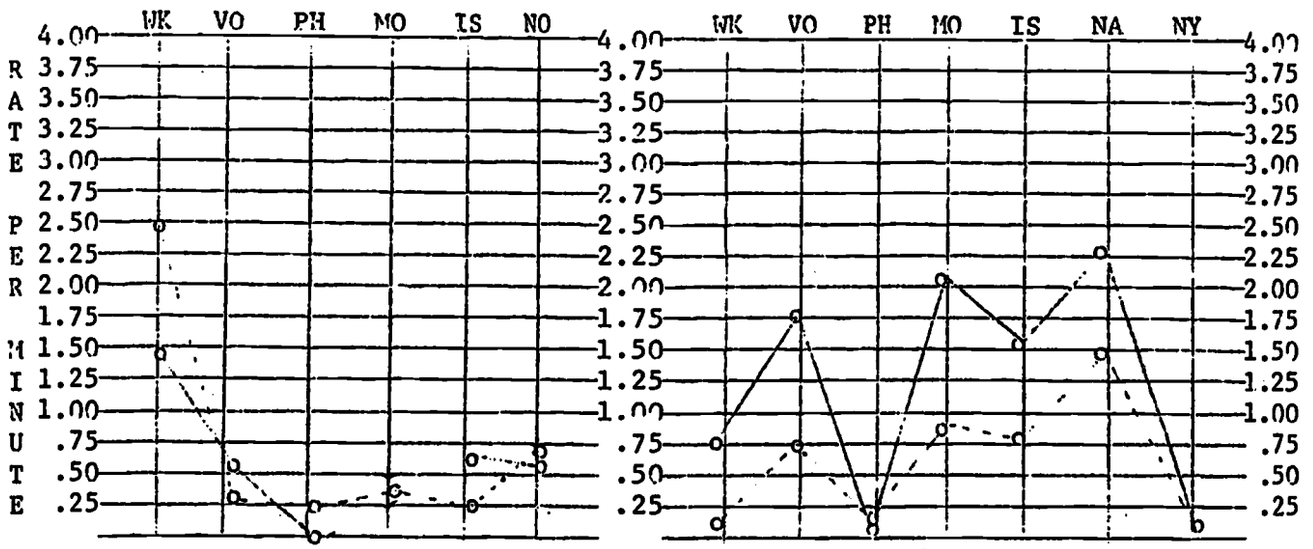
Peer

Categories

Subject IV

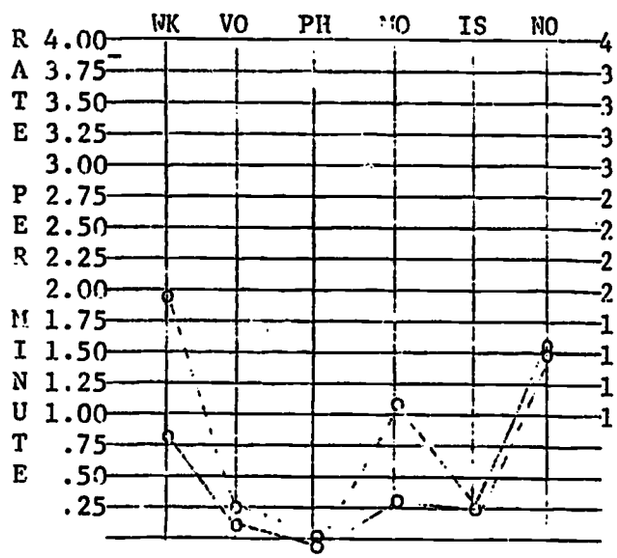
Appropriate

Inappropriate

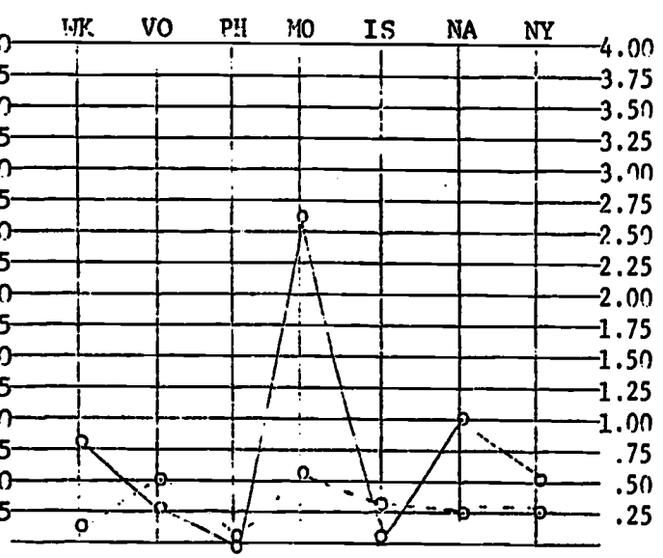


Student V

Appropriate



Inappropriate



Categories

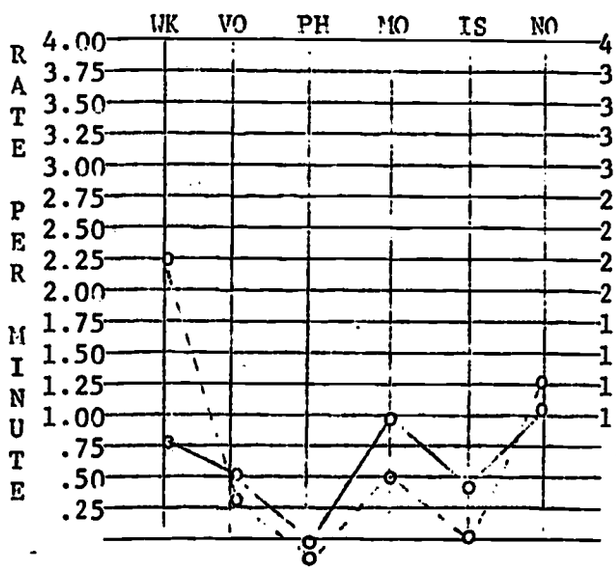
Student o——o——o

Peer o----------o----------o

Categories

Student VI

Appropriate



Inappropriate